

User Guide for **sizes**: A Small-Angle Scattering Analysis Program

Pete R. Jemian

out-of-date: 1985

Portions of this document are from the documentation supplied with the code **MAXE** (documentation by Ian Culverwell, UKAEA-Harwell, 23 February 1987) and with its modifications called **MAXE2** (modifications by Andrew Allen, UKAEA-Harwell, 19 July 1989) and **MaxSas** (modifications by Pete Jemian, 1991-1995).

1 Copyright

Copyright ©1995, Late-Nite (TM) Software

Permission to use, copy, and distribute this software and its documentation for any purpose is hereby granted without fee, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation, and that the name of Late-Nite (TM) Software not be used in advertising or publicity pertaining to distribution of the software without specific, written prior permission.

Permission to modify this software and its documentation is not granted without express, written permission for each and every instance. Late-Nite (tm) Software makes no representations about the suitability of this software for any purpose. It is provided "as is" without express or implied warranty.

Portions of this code are, or have been adapted, from the book "Numerical Recipes in C."

```
LATE-NITE (TM) SOFTWARE MAKES NO REPRESENTATIONS
ABOUT THE SUITABILITY OF THIS SOFTWARE FOR ANY
PURPOSE. IT IS PROVIDED "AS IS" WITHOUT EXPRESS OR
IMPLIED WARRANTY. LATE-NITE (TM) SOFTWARE DISCLAIMS
ALL WARRANTIES WITH REGARD TO THIS SOFTWARE,
INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY
AND FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT
SHALL LATE-NITE (TM) SOFTWARE BE LIABLE FOR ANY
SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY
DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA
OR PROFITS, WHETHER IN AN ACTION OF CONTRACT,
NEGLIGENCE OR OTHER TORTUOUS ACTION, ARISING OUT OF
OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS
SOFTWARE.
```

If you wish to make improvements to this code, please contact the author at jemian@anl.gov and you will be counted among the blessed whose camels spit dates. Those who wish to violate this copyright, or its spirit, will be considered lower than the ancestors of the parasites infesting the toes of dung flies and will be treated appropriately.

Contents

1 Copyright	1
2 Introduction	3

3	Theory	3
3.1	Small-Angle Scattering	3
3.1.1	Wavelength Smearing	3
3.1.2	Slit Smearing	3
3.2	Form Factor Choices	3
3.2.1	spheres	3
3.2.2	ellipsoids	3
3.2.3	globules	3
3.2.4	rod globules	3
3.2.5	disk globules	3
3.3	Lagrange Undetermined Multipliers	3
3.3.1	Maximum Entropy	3
3.3.2	Regularization	3
4	Operation	3
4.1	input	3
4.2	editing	4
4.3	preparation	4
4.4	solution	4
4.5	reporting	4
5	Examples	4
5.1	Synthetic Bimodal $f(D)$ Distribution	4
5.2	Synthetic Broad Distribution	4
5.3	Dow 4600 Å Polystyrene Spheres	5
5.3.1	Slit-smearred SAXS data	5
5.3.2	Desmeared SAXS data	5
6	File Formats	5
6.1	Input Command File	6
6.2	Input Data File	8
6.3	Output Fit File	8
6.4	Output Size Distribution File	9
6.5	Output Log File	10
6.6	Interim Shape Function Files	10
7	Strategies	12
7.1	The New Sample	12
7.2	Q -range	12
7.3	The <code>err</code> Parameter	12
7.4	Background	13
7.5	Linear- <i>vs.</i> Log-spaced D Bins	13
7.6	The <code>defaultDistLevel</code> Parameter	13
7.7	The <code>iterMax</code> Parameter	13
8	References	14

2 Introduction

3 Theory

3.1 Small-Angle Scattering

3.1.1 Wavelength Smearing

3.1.2 Slit Smearing

3.2 Form Factor Choices

3.2.1 spheres

3.2.2 ellipsoids

3.2.3 globules

The model for globules was developed by AJ Allen.

3.2.4 rod globules

The model for rod-shaped globules was developed by AJ Allen.

3.2.5 disk globules

The model for disk-shaped globules was developed by AJ Allen.

3.3 Lagrange Undetermined Multipliers

3.3.1 Maximum Entropy

3.3.2 Regularization

4 Operation

The program is divided into several distinct phases: input, editing, preparation, solution, and reporting.

4.1 input

First, the program reads the input command file and stores all the parameters. As a cross-check for the investigator, the program then prints out all the parameters it has interpreted from the input command file. If you are having trouble with the program, check this output first, to see if it matches what you have told the program to do.

Memory space is then allocated for the output size distribution. Next the program reads the input data file in two passes. The first pass is to count the number of data lines in the file. After the first pass, memory is allocated for the arrays necessary to hold the input data and other information required in the analysis. A second pass through the data file is used to read the input data.

4.2 editing

After the input data has been read, the input data is edited to retain only the data points within the specified Q range and to convert the data to units of absolute cross-section per unit solid angle per steradian ($I_j, \text{cm}^{-1} = \text{fac} \times (i_j - \text{bkg})$).

4.3 preparation

Preparation involves the calculation of the *grid* matrix for each combination of Q_j and r_i . This step can take some time for some shape models, such as ellipsoids of revolution measured with both wavelength-smearing and slit-smearing.

Presently there is only one choice of shape function (ellipsoids of revolution with possible wavelength and slit-length smearing) but others are planned. The investigator selects which shape function with the `shapeModel` parameter in the input data file.

4.4 solution

After the preceding steps are complete, the matrix problem can be submitted for solution. The investigator selects which analysis method to use via the `analysisType` parameter in the input command file.

Each of the analysis methods will provide periodic updates as to their progress. It is possible that an analysis method may not converge to a solution. This is more likely with the MaxEnt method. In that case, the program will not continue to the *reporting* step.

4.5 reporting

After a successful completion of the *analysis* step, the program will report the output size distribution and fit to the input data. Finally, the program will quit. (Note: on the Macintosh version of the program, the program will wait for the investigator to press [return] before quitting.)

5 Examples

There are several example files included with the distribution to aid in the use of the program. These are:

- `model2.cmd`: Analysis of a synthesized broad size distribution.
- `ps-dsm.cmd`: Analysis of 460 nm polystyrene spheres with a narrow distribution.
- `ps-smr.cmd`: Analysis of 460 nm polystyrene spheres with a narrow distribution when the input data has not been corrected for slit-smearing collimation effects.
- `test.cmd`: Analysis of a synthesized bimodal size distribution.

5.1 Synthetic Bimodal $f(D)$ Distribution

`test.cmd`

5.2 Synthetic Broad Distribution

`model2.cmd`

Figure 1: Small-angle scattering data from Dow polystyrene spheres. Analysis method accounts for the slit-smearing of the data. \circ : original data; $---$: fitted data

Figure 2: Size distribution of Dow polystyrene spheres calculated by the regularization method. 4600 Å was the reported size.

5.3 Dow 4600 Å Polystyrene Spheres

A sample of polystyrene spheres from Dow (batch xxxxx), reported diameter 4600 Å, was prepared by placing several drops of the spheres, suspended in liquid, onto scotch tape on a sample holder and then drying the liquid off. The resultant sample was a densely packed cake of the spheres. The scattering was measured on the double-crystal diffractometer at beam line X23A3 of the National Synchrotron Light Source, Brookhaven National Laboratory. Since the thickness of the sample was indeterminate, it was not possible to place the scattered intensity from this sample on a scale of absolute cross-section per unit volume per steradian.

5.3.1 Slit-smearred SAXS data

After removal of the measured underlying rocking curve of the Si (111) optics, the data was analyzed by the program `sizes` for a distribution of scatterers using the regularization method giving `ps-smr.cmd` as the input command file. The data recorded by the double-crystal diffractometer has been modified by the effect of slit-length collimation smearing, an effect which tends to smooth out the regular oscillations in the scattered intensity that arise from a low polydispersity in the size distribution.

The MaxEnt method has difficulty reconstructing the size distribution from this data set. The oscillations are of small amplitude on the data set so that the MaxEnt procedure overshoots the solution and does not appear to recover. This could be due to the specific MaxEnt search engine used. A replacement of this engine may improve this method. However, the regularization method is capable of producing a solution that both fits the data and seems credible with what is already known about the sample. The fit to the smeared data is shown in Fig. 1. There is still some systematic difference in the fit, probably due to a distribution of sizes not probed in the analysis. The volume fraction size distribution, $f(D)$, is shown in Fig. 2. It appears that the peak in the size distribution is at 4770 Å.

5.3.2 Desmeared SAXS data

`ps-dsm.cmd`

6 File Formats

All files used by the program, both input and output, are in the *text* format native to the computer system being used. For most computers, that format is ASCII. This keeps the files transportable between systems and easily inspected by the investigator.

Be certain, when modifying a file with a word processing program that “text” or “text with line breaks” is chosen when writing the file.

Many of the files allow the use of comment lines. In most cases, these comment lines begin with a # character. The correct form of comment line is described with each type of file below.

```

test : Project Name (only 1st item is read)
test.sas : SAS file, contains columns: Q i esd
1e-08    100 : qMin qMax, 1/A (1.0e-8 to 100 means all data)
        100 : rhosq : scattering contrast, 10^20 1/cm^-4
        1 : fac   : I = fac * ( i - bkg )
        1 : err   : ESD = fac * err * esd
        0.1 : bkg  : I = fac * ( i - bkg )
        1 : shapeModel (1=spheroids, no others yet)
        1 : Aspect Ratio
        0 : slitLength, 1/A
        1 : Bin Type (1=Lin, 0=Log)
        40 : nRadii
        25    900 : dMin dMax, A
        1 : n, in N(D)*V^n
        1.0e-6 : defaultDistLevel (MaxEnt only)
        32 : IterMax
        10 : iPlot (ignored for now)
        0.0002 : dLambda/Lambda
        0 : analysisType (1=MaxEnt, 0=regularization)

```

Figure 3: `test.cmd`: Example of an input command file.

6.1 Input Command File

The program uses an input command file to select the various parameters to be used in the analysis. The input command file is also useful as a reference for the documentation of a particular analysis. The name of the input command file is passed to the program on the command line (or pseudo-command line window on the Macintosh version of the program). For the input command file shown (`test.cmd`), that line would look like:

```
sizes test.cmd
```

Note: The input command file can have any name. An example of an input command file is shown in Fig. 3. Use of the extension `.cmd` is purely arbitrary.

In general, each line of the input command file contains two types of information: parameter(s) used by the program and comments for the investigator. The first one or two fields on each line are for the parameters, the rest of the line is ignored by the program. White space (tabs or spaces) is used to separate each field. The lines of the input command file must appear in the order shown or the program will perform with unpredictable results, usually crashing in the process.

Project name This parameter will be used to name the output files of the program. The output files have the form of `<project>.<ext>`. The program will assign the file extensions. (Investigators with DOS-style 8.3 file names can maintain compatibility by choosing project names with 8 or fewer characters.) It is possible to prepend a path name onto the project name if output is to be located on another path. An example of this on a UNIX system might be: `../results/test`.

SAS file This is the name of the input data file. If the file is not in the same directory or folder as the input command file, then this must be either a link (or alias) to the file or include the path name to the file. An example of this on a UNIX system might be: `../data/test.sas`.

qMin qMax (\AA^{-1}) Analyze all data points from the input data file between qMin and qMax. (1.0e-8 to 100 means “analyze all data with positive Q .”)

rhosq (10^{20} cm^{-4}) Scattering contrast $|\Delta\rho|^2$; difference in the scattering length densities of the scatterer and its surroundings.

fac $I_j = \text{fac} \times (i_j - \text{bkg})$, Factor used to scale the intensities from the input data file to absolute cross section (cm^2) per unit volume (cm^3) per steradian (sr).

err $\sigma_j = \text{fac} \times \text{err} \times \text{esd}_j$ Additional factor used to scale the errors from the input data file to absolute cross section (cm^2) per unit volume (cm^3) per steradian (sr).

bkg $I_j = \text{fac} \times (i_j - \text{bkg})$, Constant background to be subtracted from the intensities from the input data file.

shapeModel Selects which shape function to use. (1=spheroids, no others yet)

Aspect Ratio Aspect ratio, β , of the prototypical scatterer of radii: $r \times r \times \beta r$

slitLength ($l_0, \text{\AA}^{-1}$) For input data that is affected by slit smearing, this is the slit length to use in the analysis. The slit-length distribution function is a rectangular profile. If there is no slit-length smearing, set this parameter to 0. (No slit-smearing will be calculated when $l_0 < Q_1$.)

Bin Type Selects whether the bins in the size distribution will be evenly spaced on a linear (1) or a logarithmic (2) scale.

nRadii Number of bins in the size distribution.

dMin dMax (\AA) Minimum and maximum diameters in the size distribution.

n Volume weighting exponent of the size distribution, as in $N(D) V^n(D)$. $n = 0$ for a number distribution, $n = 1$ for a volume fraction distribution, or $n = 2$ for an “intensity” distribution. Use of values other than those indicated here, such as non-integer values, is not recommended. The output of the program is likely to be ill-behaved in such circumstances.

defaultDistLevel (MaxEnt only) The MaxEnt analysis method requires the investigator to make an initial estimate of the magnitude of the size distribution below which the result is meaningless. This estimate is in terms of amount, not in terms of size.

IterMax Maximum number of iterations allowed before the program will give up and report “no solution.”

iPlot (ignored for now)

dLambda/Lambda $\Delta\lambda/\lambda$. Half-width of incident wavelength distribution. Presently, the wavelength distribution is a triangular function. This can be easily changed to a Gaussian function by editing the compiler variable WL_DIST_TYPE in file `grid.h` and recompiling the code. For a Gaussian distribution, $\Delta\lambda/\lambda$ is the σ of the distribution.

analysisType Selects the analysis method to be used. Presently, the choices are regularization(0) or Max-Ent (1).

6.2 Input Data File

The input data file contains the data to be analyzed. The format of the input data file allows for *comment lines* and *data lines* in the file.

A comment line is any line that begins with a “#.” Any comment lines in the input data file will be written to the output fit and size distribution files. Also to the output log file, if it is generated. It is possible to remove individual data points from the input data file by commenting them out. In the output files, these removed data points will appear with the other program comments. Up to 4000 characters in all comment lines will be retained by the program.

The j -th data line consists of three numbers, all on the same line. These numbers should be Q_j I_j σ_j , where Q_j , the magnitude of the j -th scattering vector, is in units of \AA^{-1} . I_j and σ_j are the j -th scattered intensity and estimated standard deviation of I_j , respectively. The program has an internal conversion routine to convert these into units of absolute cross-section per unit volume cm^{-1} if needed. The numbers on the data lines can be separated by either tabs or spaces. However, all three numbers must be on the same line since the program counts the number of data lines in order to determine how many data points are in the file.

The Q_j do not need to be regularly spaced. They do, however, need to be positive. The I_j do not need to be all positive, even after the removal of the background value. The program will almost certainly have difficulty if many of the I_j are negative.

Here’s an example of the first few lines of an input data file:

```
# test.sas
#
# Q,1/A      SAS,1/cm    esd,1/cm
4.0157139E-03 3497.473 90.72816
4.5408653E-03 3340.003 84.95314
5.0095972E-03 3322.474 79.63133
...
```

6.3 Output Fit File

Once the program has determined that the convergence criteria have been met, then a data file containing the fit to the selected size distribution will be created. If a file with the name <project>.fit already exists, it will be overwritten with the new fit.

The first line in the output fit file describes the project name, the analysis method, and the time/date stamp of the completion of the analysis. It is written to the file as a comment line with the format:

```
fprintf (path,
        "# project: %s, %s analyzed: %s\n",
        ProjectName, analysisMethod, timeDate);
```

If there were any comment lines in the input data file, they will be copied to the output fit file as comments. The data is then written in the file with the format:

```
fprintf (path,
        "%lg\t%lg\t%lg\t%lg\t%lf\n",
        Q[j], SAS[j], dSAS[j], fit[j], (SAS[j]-fit[j])/dSAS[j]
        );
```

The format means that the fields are separated by “tabs,” a format accepted as the field delimiter by many plotting and spreadsheet programs. These variables have the meaning:

Table 1: Possible names for the output size distribution file, dependent on the input parameter n .

n	$N(D) \times V^n(D)$	file name	description
0	$N(D)$	<project>.N-D	number distribution
1	$f(D)$	<project>.f-D	volume fraction distribution
2	$i(D)$	<project>.i-D	intensity distribution

- `Q[j]`: The independent variables, Q_j , from the input data file
- `SAS[j]`: The small-angle scattering intensity, $I(Q_j)$ (1/cm) calculated from the input data and the applied scaling factor and background. $I_j = \text{fac} \times (i_j - \text{bkg})$ is the conversion function.
- `dSAS[j]`: The estimated standard deviation of the small-angle scattering intensity, σ_j (1/cm) calculated from the input data and the applied scaling factor. $\sigma_j = \text{fac} \times \text{err} \times \text{esd}_j$ is the conversion function.
- `fit[j]`: The intensity (\hat{I}_j) re-calculated from the fitted size distribution.
- `(SAS[j]-fit[j])/dSAS[j]`: The standardized residuals ($z_j \equiv (I_j - \hat{I}_j)/\sigma_j$) of the fit. For a good fit, z_j should be scattered randomly between -1 and $+1$.

All comment lines in the output fit file have a “#” as the first character. All other lines in the file are data lines.

Here’s an example of the first few lines of an output fit file:

```
# project: test, regularization analyzed: 95.11.02, 07:53:49
# data file comment:
# test.sas
#
# Q,1/A          SAS,1/cm    esd,1/cm
#
# Q,1/A SAS,1/cm dSAS,1/cm fit,1/cm z
0.00401571 3497.37 90.7282 3610.6 -1.247956
0.00454087 3339.9 84.9531 3392.34 -0.617278
0.0050096 3322.37 79.6313 3189.1 1.673591
...
```

6.4 Output Size Distribution File

The output size distribution file may take one of three names depending on the input parameter n , which weights the volume term in the size distribution ($N(D) \times V^n(D)$).

Table 1 shows the various possibilities. The most common weighting is the volume fraction distribution, $f(D)$, where $n = 1$. The number distribution, $N(D)$, is used to make comparisons with other particle counting methods, such as transmission electron microscopy. The intensity distribution, $i(D)$, is rarely used.

Here’s an example of the first few lines of an output size distribution file:

```

# project: test, analyzed: 95.11.02, 07:53:49
# data file comment:
# test.sas
#
# Q,1/A          SAS,1/cm    esd,1/cm
#
# D,A f,1/A
25 3.6243e-06
47.4359 2.95622e-06
69.8718 -5.37022e-06
...

```

6.5 Output Log File

A log file can be obtained for some versions of the program. On the Macintosh, a special feature of the Symantec C compiler is used to simultaneously log any screen output into a file. The name of the file so produced is `<project>.log`. On UNIX systems, the same log file can be obtained by directing the program output into a file with a UNIX command such as

```
sizes input.cmd > project.log
```

where `input.cmd` is the name of the input command file and `project.log` is the name of the output log file. However, use of this UNIX command will redirect the output into the file; no program output will appear on screen. Fig. 4 gives an example of an output log file.

6.6 Interim Shape Function Files

Files containing the calculated shape functions, $S(u)$, can be obtained for some of the chosen shape functions.

For the shape function modeling ellipsoids, the interim shape function files may be obtained by defining the compiler variable:

```
#define WRITE_INTERMEDIATES    /* useful when debugging */
```

in the file `grid.h` and rebuilding the executable file.

The data lines of the file are written with the format:

```
fprintf (path, "%lg\t%lg\n", x[i], y[i]);
```

in the routine: `WriteFile`.

The files written by the ellipsoid model are:

`P_lambda.dat` : wavelength distribution $p(\lambda)$

`wave-i.dat` : i -th wavelength smearing integrand $p(\lambda) S_\lambda(u_i \bar{\lambda}/\lambda)$

`S_s.dat` : Form factor for spheres $S_w(u)$

`S_e.dat` : Form factor for ellipsoids $S_e(u)$, if applicable

`S_w.dat` : Wavelength-smearred form factor for ellipsoids $S_{e,\lambda}(u)$, if applicable

```

General Size Distribution Analysis of SAS data: 95.11.02, 07:53:40

Review of input parameters for analysis:
      test : ProjectName
      test.sas : SASfile
1e-08      100 : qMin qMax, 1/A
           100 : RhoSq
           1 : fac
           1 : err
           0.1 : Bkg
           1 : shapeModel
           1 : Aspect
           0 : slitLength, 1/A
           1 : Bin type (1=Lin, 0=Log)
           40 : num bins
25        900 : Dmin & Dmax
           1 : n, in N(D)*V^n
1e-06     32 : defaultDistLevel
           32 : IterMax
           10 : iPlot (ignored for now)
0.0002    0 : dE_E
           0 : analysisType (0=regularization, 1=MaxEnt)

Reading SAS from file: test.sas ...91 points were read.
91 points remained after Q-range editing.
file comments:
# test.sas
#
# Q,1/A      SAS,1/cm    esd,1/cm
Preparing the G(h,r) matrix ... (95.11.02, 07:53:42)
Beginning the solution routine ... (95.11.02, 07:53:44)
      log10(a)      ChiSqr      S      Q=ChiSqr + a*S
           0          59.4836      0.00410335      59.4877
           50         78936.3      1.36203e-81      78936.3
           25         78936.3      1.36203e-31      78936.3
           12.5       10313.3      2.04245e-09      16772.1
           6.25       79.9625      7.13782e-06      92.6555
           9.375      1205.18      1.96006e-07      1669.98
           7.8125     153.436      2.96259e-06      345.821
           7.03125    89.9127      4.99894e-06      143.632

```

Figure 4: test.log: Example output log file.

7 Strategies

Sometimes, it is helpful to have a little coaching. . . There are many parameters that can be adjusted in the input command file. Complementary investigations or other *a priori* knowledge about the sample can be beneficial in expediting the analysis. This section will attempt to describe how to use the various parameters to advantage while homing in on an acceptable result.

It is unlikely that subtle changes in the specified $\Delta\lambda/\lambda$ will affect the result unless the polydispersity of the sample is quite low (*i.e.*, the size distribution is sharply peaked).

The `rhosq` and `fac` parameters will only affect the magnitude of the output size distribution and will not affect the shape of the distribution or the convergence of the chosen analytical method.

7.1 The New Sample

For the first-time analysis of SAS data from a sample, one may not be certain of the best values for all of the parameters. Experience is a good guide. So is *a priori* knowledge or complementary investigations of the sample.

The convergence tests of the *MaxEnt* analysis are more stringent than for the *regularization* method. While the latter allows the resultant size distribution to have negative values, the regularization result obtained can be used as a guide to better estimate some of the input parameters.

In most cases, there is no penalty for increasing the range of D and the number of points in the distribution. Be warned that increasing the number of points will increase the computation time proportionately. However, an inadequate number of points may fail to reproduce the measured scattering correctly, even if the other parameters are correct. (Verify this on the `test` data set by setting the number of D to 20 or less.)

If the chosen D range spans more than about a decade, then it is more advisable to use the `log` spacing for D_i values, since it likely *maps* more closely to the spacing of the input Q_j values.

7.2 Q -range

Choose the input Q range from the data based on your judgement. It may not be possible to analyze simultaneously 4 decades of Q data (you should be so fortunate to have this type of data). Some of the smallest Q values may be affected by penumbra (*shadowing*) of the beam stop.

The program will perform unpredictably and likely crash for data with $Q_j \leq 0$. Only use positive Q_j -values!

Don't waste time including a lot of flat background or negative intensity points in the analysis. This does not improve the result.

7.3 The `err` Parameter

The `err` parameter is the most easily abused parameter. One is tempted to increase it slightly in order to force convergence of a particular analysis. The purpose of `err` is to correct for inadequately estimated errors in the input data. Remember that `err` is applied as a constant factor to *all* the input `esdj`.

Use caution! Be sure that the final value of `err` chosen has physical significance. When tracking down better estimates of the other parameters, adjusting `err` (usually by $\times 3$ to $\times 10$) can be useful to force convergence. Remembering that one convergence criterion for MaxEnt and for regularization is that $\chi^2 = M$, where M is the number of input data points, one can force convergence to a different point in the entropy or smoothness of the size distribution. Just make sure that the σ_j in the final analysis have physical significance.

7.4 Background

In some input data sets, the sample background may not be a constant value or it may not be known at all. If it is not flat, investigate the reasons for this! Is it due to another source of small-angle scattering, such as clustering? Is it due to an overlap in the Q -range with the interatomic structure factor? Is the shaped-background function due to an instrumental effect? All of these possibilities should be handled *before* the data is analyzed with this program or artifacts will certainly appear in the resultant size distribution.

If the background is not known, several methods exist to estimate it. One, which only applies when a well-established Porod region exists, is the *Porod* plot wherein the data is plotted as IQ^4 vs. Q^4 (Q^3 for slit-smear data). According to Porod (reference Porod in Glatter and Kratky here), the slope of the plot, ignoring the first few data points, is the instrumental background.

Often, the background can be estimated visually by examination of a plot of $\log(I_j)$ vs. $\log(Q_j)$. At the highest Q_j values, the data may begin to go flat. Use your judgement.

7.5 Linear- vs. Log-spaced D Bins

As stated above, the log-spaced D bins will likely *map* more evenly onto the input Q_j values if the ratio $D_{max}/D_{min} > 30$ or so. This is because the scattering at higher angles tends to over-sample the smaller size scatterers and under-sample the larger sizes because most step-scanning (or radial binning) algorithms tend to use constant ΔQ intervals. The log spacing provides more D bins at the smaller sizes and fewer at larger sizes.

Do not become very exercised about this difference as it is truly second order for the most common input data sets, where $Q_{max}/Q_{min} < 100$. Also, it may not be important for most size distributions that are clearly peaked within the size range investigated. However, for the SAS data arising from a very broad size distribution (nearly fractal, some may say), the log-spacing is the correct choice.

7.6 The defaultDistLevel Parameter

For a better understanding of the purpose of the the `defaultDistLevel` parameter used in the `MaxEnt` analysis method, it is best to read Skilling and Bryan's paper. If the order of magnitude of the value is not known, switch to the regularization method and obtain an output size distribution. Make your first estimate of `defaultDistLevel` as 5% (or so) of the peak of the distribution. Switch back to the `MaxEnt` method and proceed. Convergence will probably be obtained. Observe that by adjusting `defaultDistLevel`, the structure of the distribution at (usually) the smaller sizes will be affected most.

This procedure is best to follow when switching the volume weighting parameter, n , as in $N(D) \times V^n(D)$ since the magnitudes of $N(D)$, $f(D)$, and $i(D)$ are very different.

7.7 The iterMax Parameter

Increasing the `iterMax` parameter is useful for some troublesome analyses. For most analyses, convergence should be obtained in 15-25 iterations of the selected method. For `regularization`, the number of iterations is directly proportional to the number of significant bits of precision for a double-precision floating point number. This is because the search for the Lagrangian parameter α is done by bisection. Usually 32 is a reasonable number, even though the precision of a `double` may be 64 or greater.

Troublesome analyses occur more often with `MaxEnt` analyses, when the $\chi^2 = M$ criterion has been met but the parallelism criterion between the $\vec{\chi}^2$ and \vec{s} vectors has not been reached. At this point, the `MaxEnt` search engine is shifting values in the distribution between D_i bins to make these two vectors parallel within 5% (compiler variable `TEST_LIMIT` in file `maxent.c`).

Other troublesome analyses happen when the input data has some small structure with respect to the σ_j . These are just the hard facts of life; you may need better data.

8 References

1. J. Skilling and R.K. Bryan; *Mon Not R Astr Soc* **211** (1984) 111 - 124.
2. J.A. Potton, G.J. Daniell, and B.D. Rainford; Proc. Workshop on Neutron Scattering Data Analysis, Rutherford Appleton Laboratory, UK, 1986; ed. M.W. Johnson, *IOP Conference Series* **81** (1986) 81 - 86, Institute of Physics, Bristol, UK.
3. I.D. Culverwell and G.P. Clarke; Proc. Workshop on Neutron Scattering Data Analysis, Rutherford Appleton Laboratory, UK, 1986; ed. M.W. Johnson, *IOP Conference Series* **81** (1986) 87 - 96, Institute of Physics, Bristol, UK.
4. J.A. Potton, G.J. Daniell, & B.D. Rainford; *J Appl Cryst* **21** (1988a) 663 - 668.
5. J.A. Potton, G.J. Daniell, & B.D. Rainford; *J Appl Cryst* **21** (1988b) 891 - 897.
6. L.C. Roess & C.G. Shull; *J Appl Phys* **18** (1947) 308-313.